Journal of Informetrics 13 (2019) 485–499

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Regular article

Predicting citation counts based on deep neural network learning techniques

Ali Abrishami, Sadegh Aliakbary*

Faculty of Computer Science and Engineering, Shahid Beheshti University G.C., Tehran, Iran

ARTICLE INFO

Article history:
Received 2 October 2018
Received in revised form 4 February 2019
Accepted 23 February 2019
Available online 11 March 2019

Keywords:
Informetrics
Citation count prediction
Neural networks
Deep learning
Scientific impact
Time series prediction

ABSTRACT

With the growing number of published scientific papers world-wide, the need to evaluation and quality assessment methods for research papers is increasing. Scientific fields such as scientometrics, informetrics, and bibliometrics establish quantified analysis methods and measurements for evaluating scientific papers. In this area, an important problem is to predict the future influence of a published paper. Particularly, early discrimination between influential papers and insignificant papers may find important applications. In this regard, one of the most important metrics is the number of citations to the paper, since this metric is widely utilized in the evaluation of scientific publications and moreover, it serves as the basis for many other metrics such as h-index. In this paper, we propose a novel method for predicting long-term citations of a paper based on the number of its citations in the first few years after publication. In order to train a citation count prediction model, we employed artificial neural network which is a powerful machine learning tool with recently growing applications in many domains including image and text processing. The empirical experiments show that our proposed method outperforms state-of-the-art methods with respect to the prediction accuracy in both yearly and total prediction of the number of citations.

1. Introduction

Nowadays, many researchers are working on scientific projects world-wide and writing research papers. As a result, many papers are being published everyday with different scientific qualities and impacts. Therefore, the need to evaluating published papers and assessing their quality is overwhelming. A variety of criteria exist in the literature for evaluating the quality of a scientific paper, but one of the most important evaluation metrics is the number of citations to the considered paper. The citation count is a significant indicator since it is widely used for measuring the impact of a paper (Eugene, 1998; Moed, 2005; Oppenheim, 1995) and moreover, it has been used as the basis for many other metrics such as h-index (Hirsch, 2005), impact factor (Eugene, 2006), i-10 index, and other evaluation metrics for journals, conferences, researchers, and research institutes (Moed et al., 2012; Wildgaard, Schneider, & Larsen, 2014).

We consider the problem of predicting the citation count of a scientific paper. This problem has many applications in different domains. With the increasing amount of published papers, researchers need to recognize the more influential papers in advance, so that they can plan their research direction (Amjad et al., 2017; Yan, Huang, Tang, Zhang, & Li, 2012). Moreover, by predicting the citation count of a paper, we can evaluate the future impact of the paper authors, with potential

* Corresponding author.
E-mail addresses: a.abrishami@mail.sbu.ac.ir (A. Abrishami), s_aliakbary@sbu.ac.ir (S. Aliakbary).

https://doi.org/10.1016/j.joi.2019.02.011

applications in hiring researchers and faculties, and granting awards and funds. Various efforts exist in the literature for gaining such insights about the future impact of researchers (Chan, Mixon, & Torgler, 2018; Fiala & Tutoky, 2017; Havemann & Larsen, 2015; Revesz, 2014).

In this paper, we propose a method for predicting the citation count of a scientific paper based on its citations during the early years of publication. In other words, the proposed method takes the citation count of a paper in a few years after its publication (namely 3–5 years), and then predicts its citations in a more long-term period (e.g. from 5th to 15th year after publication). In this problem, we do not investigate other sources of information such as author features, journal properties, and the content (text) of the paper. Although some existing works include more sources of information as the inputs of the citation prediction method, we limit the inputs as simple as the citation pattern of the early publication year, in order to keep the problem definition simple, general, and also applicable in other domains. This problem has already gained attention in the literature, and various methods are proposed to solve this problem (Cao, Chen, & Ray Liu, 2016; Wang, Song, & Barabási, 2013).

It is worth noting that citation prediction is not a trivial task, since scientific papers show different patterns of citations. For example, some papers remain unnoticed for many years and then, attract a lot of attention (this phenomenon is called "sleeping beauty in science") (Ke, Ferrara, Radicchi, & Flammini, 2015; Li & Ye, 2016; van Raan, 2004). Some other publications gradually gain less citations due to emergence of new methods or losing applicability. Consequently, a single rule or a simple model cannot effectively predict the future citations of a paper, and more powerful methods are needed.

In the terminology of machine learning methods, we considered the citation prediction as a regression learning problem, and then we utilized artificial neural networks as a powerful model for learning the prediction task. Based on the citation patterns of many existing scientific papers, the proposed neural network is trained to predict the citation count of papers in the future. Artificial neural networks are inspired by the human brain networks, and has found many successful applications in regression and classification learning. In recent years, neural networks (particularly deep networks) have been effectively applied in various problems such as voice recognition (Dahl, Yu, Deng, & Acero, 2012; Hinton et al., 2012), object recognition (Krizhevsky, Sutskever, & Hinton, 2012; Schmidhuber, 2012), image processing (He, Zhang, Ren, & Sun, 2016; Wan et al., 2014), and text processing (Severyn & Moschitti, 2015a; Sutskever, Martens, & Hinton, 2011). We designed a customized recurrent neural network which is appropriate for learning the sequence of the citations. We have also run comprehensive experiments to show the effectiveness of the proposed method over the existing state-of-the-art baseline methods.

The rest of this paper is organized as follows: In Section 2 we review the related work. In section 3 we formulate the problem and present the proposed method. Section 4 describes the prepared dataset which is utilized in our experiments. In Section 5 we evaluate the proposed method and compare it with the state-of-the-art baselines. Finally, we conclude the paper in Section 6 and describe the future works.

2. Related work

Many studies exist in the literature for predicting the impact and success of scientific works. The existing works aim different goals such as citation count prediction for scientific papers (Bornmann, Leydesdorff, & Wang, 2014; Cao et al., 2016; Castillo, Donato, & Gionis, 2007; Dong, Johnson, & Chawla, 2015, 2016; Lamb, Gilbert, & Ford, 2018; Mazloumian, 2012; Pobiedina & Ichise, 2016; Wang et al., 2013; Yan et al., 2012; Yan, Tang, Liu, Shan, & Li, 2011; Yu, Yu, Li, & Wang, 2014), predicting highly cited papers (McNamara, Wong, Christen, & Ng, 2013; Newman, 2014; Sarigöl, Pfitzner, Scholtes, Garas, & Schweitzer, 2014), predicting h-index of the researchers (Acuna, Allesina, & Kording, 2012; Ayaz, Masood, & Islam, 2018; Dong et al., 2016), and predicting the impact factor of scientific journals (Ketcham, 2007; Silva, 2016; Wu, Fu, & Rousseau, 2008).

We may categorize the existing works based on their utilized sources of information for scientific impact prediction. In the first category, the graph of the scientific papers is used as the main source of information (Bütün, Kaya, & Alhajj, 2017; Daud et al., 2017; Klimek, Jovanovic, Egloff, & Schneider, 2016; McNamara et al., 2013; Pobiedina & Ichise, 2016; Sarigöl et al., 2014). Sarigol et al. use the co-authorship network of the scientists and the author centrality measures for predicting highly cited papers (Sarigöl et al., 2014). Many existing works tackle this problem as a link prediction problem in the citation network of the papers (Bütün et al., 2017; Daud et al., 2017; Pobiedina & Ichise, 2016). McNamara et al. also investigate the paper neighborhood properties in the citation network in order to forecast the highly cited papers (McNamara et al., 2013). Klimek et al. construct a bipartite network of papers and words (only the words in the papers abstract), and analyze this network to find the papers with the highest impact potential (Klimek et al., 2016).

The second category of the existing works utilize information that are available right after the publication of the papers. This information includes the content of the paper (paper text), the publication venue (e.g., the journal or conference), information related to the authors, the subject (research area) of the paper, and the references. Such data which are available right after the publication of a paper, may help gain insight about the impact of the paper in the future. For example, Dong et al. (2016) propose to use the following six information sources for predicting whether a paper will increase the h-index of the author in a five-year period after publication: author, topic, references, publication venue, the social network, and the temporal attributes. Mazloumian (2012) utilizes the information related to the paper authors such as published papers count, the research experience years count, the average annual citations count, and h-index, in order to predict the total author citation count in k successive years. Some methods also utilize the topic of the paper (e.g. its rank and freshness), the authors properties (e.g. rank and h-index), and the publication venue for predicting the paper citation count in the future

(Yan et al., 2011, 2012). Yu et al. utilize 24 features including information about the authors and the publication venue (Yu et al., 2014). Castillo et al. utilize information about the past publications of the authors and the co-authorship network to predict the citation count in the first few years after publication (Castillo et al., 2007). Bornmann et al. utilize the journal impact factor, number of the authors, and number of the references (Bornmann et al., 2014).

The third category of information which is utilized in existing works includes data gathered after the publication of the paper. For example, Lamb et al. (2018) investigate the popularity of a paper in the web (e.g., in the news and Wikipedia) and social networks (such as Twitter and Facebook), and then show a correlation between this popularity and the amount of citation to the paper. As the main class of interest, many existing works utilize short-term citation count of a paper for predicting its long-term citations (Cao et al., 2016; Newman, 2014; Wang et al., 2013; Yu et al., 2014). The citation count of the early years after publication is an important feature which may contribute to improve the accuracy of citation count prediction methods (Bornmann et al., 2014; Kosteas, 2018). For example, Newman (2014) predicts highly cited papers based on the short-term citation count of the papers and the computation of z-score of the citations of the papers of the same research field. In a significant research, Wang et al. designed a general relationship (formula) which predicts the citation count of a paper in the years after publication (Wang et al., 2013). In this method, the prediction formula is fitted to a specific paper by tuning the formula parameters based on the citation count of the paper in the early years of its publication. This work, which served as a leading research for citation count prediction, followed a manual designed formula which is developed by human experts. But in recent years, artificial intelligence and machine learning methods are widely utilized to replace human-inspired constant models. For example, Cao et al. predict citation count of a paper using the citation pattern of the most similar existing papers according to the citations of the early years after publication (Cao et al., 2016). This method first clusters the papers according to their similarity in short-term citation patterns, and then predicts the citation count based on the average and the centroid of the clusters. This method outperforms state of the art related works and thus is considered as one of the main baselines in our experimental evaluations.

It is worth noting that three described categories overlap, i.e., some methods belong to more than one category. For example, some existing works utilize both short-term citation counts and the author properties (Castillo et al., 2007; Yu et al., 2014) and therefore, they belong to both the second and the third illustrated categories.

In this paper, we only consider the short-term citation count (no other features) for predicting the long-term citations of scientific papers. Significant recent studies have defined the same problem statement with the same input information (i.e., only short-term citations pattern) (Cao et al., 2016; Wang et al., 2013). This is an important problem since it does not utilize extra information such as the paper text or the authors background and therefore, the results are applicable when the input data is limited.

3. Methodology

In this section, we formulate the problem and the assumptions and then, we describe our proposed method along with the details of the employed techniques.

3.1. Problem statement

Suppose that the target paper has received $c_0, c_1, \ldots, c_n$ citations respectively in the years after its publication. In other words, c_i shows the citation count of the paper in the ith year after publication. Assume that we know $c_0, c_1, \ldots, c_k$ and we want to predict $c_{k+1}, c_{k+2}, \ldots, c_n$ for a paper ($k < n$). In other words, the problem is to predict the citation count of a paper until the nth year of its publication when we already know its citation count only for the first $k+1$ years (0th to kth year) after its publication. As we described in Section 2, we only consider the citations of the first $k+1$ years as the input of the algorithm, and no other information (such as author properties or journal attributes) is utilized. The actual citation count for the ith year is called c_i, and the predicted citation count for the ith year is called $\hat{c}_i$. Moreover, we define C as the total citations of the paper from $(k+1)$th to nth year of publication, and $\hat{C}$ as the corresponding total predicted citations of the same period, which are described in Eqs. (1) and (2) respectively.

$$C = \sum_{i=k+1}^{n} c_i \tag{1}$$

$$\hat{C} = \sum_{i=k+1}^{n} \hat{c}_i \tag{2}$$

A citation prediction method is evaluated according to the accuracy of both $\hat{C}$ and $\hat{c}_i$ values. This is because an effective prediction method should estimate both yearly citations and total citations of a paper in the considered time period. Consequently, the problem is to minimize the error of $\hat{C}$ and $\hat{c}_i$ values, i.e., $\hat{c}_i$ values should be as close and correlated as possible to c_i values, and so for $\hat{C}$ and C quantities. Table 1 summarizes the defined symbols.

Table 1
Table of symbols.

c_i	Citation count of the target paper in its ith year of publication
$\hat{c}_i$	The predicted citation count of the target paper in its ith year of publication
k	The number of years after publication in which the citation counts are known
n	The number of years after publication in which the citation count should be predicted
C	Total citations of the target paper from the $(k+1)$th to nth years after its publication
$\hat{C}$	Total predicted citations of the target paper in the period of $(k+1)$th to nth years after its publication

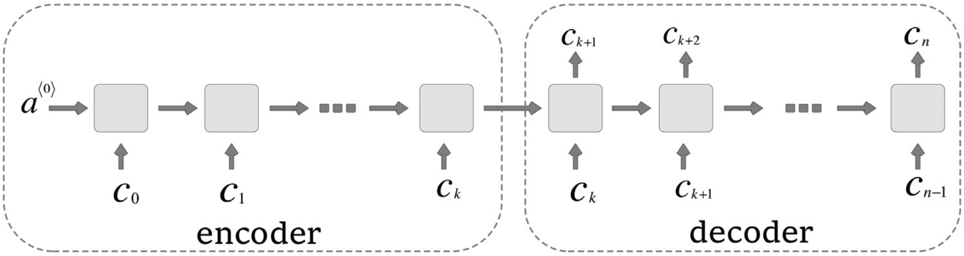


Fig. 1. The proposed neural network in the training phase.

3.2. Citation count prediction

We utilize machine learning methods in order to build a model that learns to predict citation count of a paper in the future based on its citation history. The model should predict the citation count as a non-negative integer number and therefore, the defined problem is regarded as a "regression problem" in the context of machine learning methods. Artificial neural network is one of the most powerful and effective methods for regression learning and thus, we designed a special neural network as the main component of our proposed method. A neural network is built up of several layers of neurons, and it learns to find a relationship (function) between the inputs and the outputs of the training data. In recent years, neural network has found significant applications in text processing (Lai, Xu, Liu, & Zhao, 2015; Mikolov, Karafiát, Burget, Černocký, & Khudanpur, 2010; Severyn & Moschitti, 2015b), image processing (Karpathy & Fei-Fei, 2017; Krizhevsky et al., 2012; Simonyan & Zisserman, 2014), voice processing (Graves, Mohamed, & Hinton, 2013), and many other fields. Particularly, the so called "deep learning" techniques are developed in recent years with surprising power for learning complex functions (Goodfellow, Bengio, & Courville, 2016).

A neural network is first trained in the training phase, and then it is used for prediction in the sampling phase. In our proposed method, a neural network is trained which having the citation count history of a paper in its early years of publication as the input, predicts the citation count in the future years as the output. In other words, the proposed neural network learns to predict $\hat{c}_{k+1}, \hat{c}_{k+2}, \ldots, \hat{c}_n$ values based on $c_0, c_1, \ldots, c_k$ values. The neural network is first trained using a dataset of existing papers with known citations history, and then it can be utilized as the citations estimator in the future based on the pattern of the previous citations of the considered paper. As the datasets of training and test data, we utilize many published papers with known citations history in their first n years after publication. We employ some of those papers as the training set in order to train the neural network, and then we utilize the rest of the papers as the test set in order to evaluate the accuracy of the trained neural network.

In the defined citation count prediction problem, both the inputs and the outputs form sequence of consecutive values. In such problems, Recurrent Neural Networks (RNN) (Rumelhart, Hinton, & Williams, 1986; Werbos, 1990) are known to be effective for learning the sequence of the values and therefore, we utilized RNNs in our proposed method. RNNs are capable of learning tasks in which the inputs conform an inherent sequence. For example, speech to text (Graves et al., 2013), sentiment analysis (Severyn & Moschitti, 2015b), and machine translation (Bahdanau, Cho, & Bengio, 2014; Cho et al., 2014) methods are frequently trained by RNNs since their input data are inherently sequential (rather than a set of independent features). As an alternative to RNNs, simple feedforward neural networks are unable to effectively realize the sequence nature of the input and therefore, result in less accuracy. RNNs process the input sequence in their inherent order and as the sequence is being processed, a hidden memory is built based on the so-far-visited input data and therefore, the sequence of the input is effectively considered (Chollet, 2017). In the defined citation prediction problem, the input sequence includes $k+1$ values $(c_0, \ldots, c_k)$, and the output sequence includes $n-k$ values $(\hat{c}_{k+1}, \ldots, \hat{c}_n)$. Thus, a many-to-many RNN architecture is designed in our proposed method (RNNs are classified into four categories based on the length of the input and output data: 1 – one-to-one, 2 – one-to-many, 3 – many-to-one, and 4 – many-to-many).

As the main building block in the architecture of our proposed method, we employed a deep neural network technique called sequence-to-sequence model (Sutskever, Vinyals, & Le, 2014) which has already found many successful applications in the literature (Cho et al., 2014; Sutskever et al., 2014). This technique trains models to convert sequences from one domain to sequences in another domain (e.g., to translate a French sentence to English). The model is composed of two independent neural networks called the "encoder" and the "decoder" networks. As illustrated in Fig. 1, the inputs are fed to the encoder

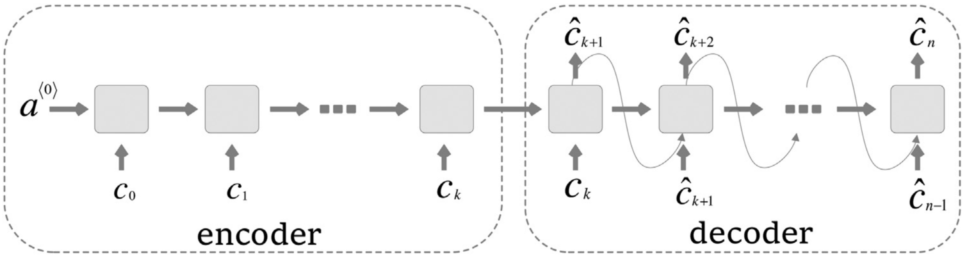


Fig. 2. The proposed neural network in the sampling phase.

and the outputs are obtained from the decoder. The encoder network aims to process the input sequence, and create its corresponding representation in another dimensional space. This representation is then forwarded to the decoder network which converts it to the output sequence. In order to predict each output neuron, the decoder utilizes all its input neurons along with the formerly predicted output neurons (see Fig. 1). In this manner, the output neurons are also fed into the network to enable learning the sequence pattern of the input data.

The proposed neural network should be trained in order to predict the citation count of the papers in the future. In the training phase, the training dataset is utilized in order to optimize the parameters of the neural network. Each tuple of the training set corresponds to the citation information of one published paper, which includes the citation count of the paper from the 0th to kth year of its publication as the input, and the actual citation count from $(k+1)$th to nth year as the output. Therefore, the inputs (citation count values for $k+1$ years) are fed to the encoder and the actual outputs (citation counts from the year $k+1$ to n) are presented to the network for training. As illustrated in Fig. 1, c_0 to c_k are the inputs of the neural network which are fed to the encoder, and c_{k+1} to c_n are its outputs which are obtained from the decoder. It is worth noting that the decoder network takes c_k as an input in the decoder network in order to predict c_{k+1}, and then c_{k+1} is needed for predicting c_{k+2} and so forth.

After the training phase, the sampling phase aims to predict the citation count for the papers which are not employed in the training phase, having the citation count of their first k years of the paper after publication. Fig. 2 illustrates the sampling phase, in which the decoder estimates the c_x for $x>k$ values. Since the neural network does not access the actual c_x data in the sampling phase, $\hat{c}_x$ predicted values are also fed to the next neuron of the decoder neural network in the sampling phase (this technique is called "teacher forcing" (Goodfellow et al., 2016)). In Figs. 1 and 2, $a^{<0>}$ illustrates the initial hidden memory (hidden state) of the encoder network. We have set this initial memory to an all-zero vector which is fed to the encoder network (this initialization is common in deep learning tasks).

The proposed method, learns to predict the future citation count of a paper based on the history of its early years citations. The employed recurrent neural network with the sequence-to-sequence model technique, enables the method to effectively learn the sequence pattern of the citations, and to predict the future citations more accurately than the baseline methods, as illustrated in Section 5.

3.3. Implementation details

In this section, we overview the implementation details of our experiments in order to make the research results reproducible. Particularly, we describe the implemented neural network and its parameters which resulted in the best outcomes based on our experiments. Such implementation details are worth noting because designing an effective deep neural network is a challenging task for this application which requires investigating many hyper-parameters and design choices.

We utilized "Keras" framework (https://keras.io) which is a well-known and widely used implementation for artificial neural networks and deep learning techniques. In addition, we used the "SimpleRNN" module of Keras for implementing the proposed neural network. In order to make the neural network capable of learning a complex function, the recurrent layers include 512 values, which means the encoder network generates a vector output with 512 dimensions. In the decoder sector, a "Dense" layer is utilized to generate the output (predicted citation count) namely $\hat{c}_i$. We did not normalize the inputs and outputs of the neural network. The "Rectified Linear Unit" (ReLU) is used as the activation function in all layers of the neural network, which is defined as $f(x)=max(0, x)$. The ReLU activation function is widely being used in recent years because it results in better precision in many applications (Glorot, Bordes, & Bengio, 2011). In order to avoid overfitting, the "Dropout" technique (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014) is used with the rate of 0.2 in the RNN layers. This technique enables more epochs (100 epochs in our experiments) in the training phase while minimizing the risk of overfitting. We implemented the RMSProp optimization algorithm, which is an effective method for training neural networks, with the learning rate of 10^{-5}. Data are fed to the network in batches 256 papers (batch-size = 256). More than half a million of parameters are trained and tuned in the training phase of the proposed method. Table 2 summarizes the implementation details of the proposed method.

Table 2
The implementation details of the proposed method.

Neural network API	Keras
RNN module	SimpleRNN
Output dimensions of the encoder	512
The output layer	Dense layer
Activation function	ReLU
Overfitting prevention technique	Dropout with 0.2 rate
Epochs	100
Optimization algorithm	RMSProp
Learning rate	10^{-5}
Batch size	256

3.4. Baseline methods

We selected three baselines in order to compare the evaluation results against. First, the "Mean of Early Years" method (MEY) is a naïve and simple prediction function which always returns the average of the known citation count in the first years of paper publication as the predicted citation count in the future. Eq. (3) shows the MEY prediction method. For many papers, the citation count in the early years of publication is similar to long-term citations. Although MEY is a simple prediction function, it shows relatively good prediction accuracy in some situations, and outperforming MEY is not a trivial task for citation prediction methods.

$$\hat{c}_{k+1} = \hat{c}_{k+2} = \cdots = \hat{c}_n = \frac{\sum_{i=0}^{k} c_i}{k+1} \tag{3}$$

The second and the third baselines are proposed by Cao et al. (2016). As illustrated in Section 2, this research has already outperformed important existing methods, such as (Wang et al., 2013), and thus we considered it as one of the main baselines in our experiments. The second baseline, which is called AVR, finds the most similar papers published in the same journal according to the citation count of the paper in the early years of publication and then, utilizes the average citations of those found papers in the subsequent years as the predicted citation count of the paper. The third baseline, which is called GMM, also finds the most similar papers published in the same journal according to the citation count of the early publication years but then, clusters the found papers in three groups using Gaussian mixture model (GMM) algorithm and then, the most similar centroid to the target paper is selected. The citation pattern of that centroid is utilized as the prediction function for the citation count of the target paper. Finally, our proposed method is called NNCP (Neural-Network-based Citations Prediction) in the evaluation reports.

It is also worth noting that AVR and GMM methods utilize an L parameter which indicates the number of nearest neighbor papers in the dataset to the target paper. In other words, AVR and GMM methods first find L papers, $y^{(1)}, y^{(2)}, \ldots, y^{(L)}$, from the database of past papers D with smallest matching error $e_x(y)$, where x is the target paper which we want to predict its citations. In order to optimize L parameter for the baseline methods, we considered the papers published in Nature, and using cross-validation technique, we selected $L = 20$ which resulted in the best average accuracy in AVR and GMM methods for $L < 100$ values.

4. Data

We extracted a dataset of published papers along with their citations from the Web of science (WoS) citation database which is a well-known online scientific citation indexing service. In this dataset, we considered the publications of five prestigious journals: Nature, Science, NEJM (The New England Journal of Medicine), Cell and PNAS (Proceedings of the National Academy of Sciences). The dataset covers the papers which are published from 1980 and before 2003, and it includes 14 years of the citations for each paper (all citations before 2017). The mentioned journals are high-impact publications with a long history. Moreover, one of our main baseline methods (Cao et al., 2016) also included similar journals and therefore, we considered such journals for fair comparison of evaluations.

In order to separate training and test data, we considered the published papers between 1980 and 1997 as the training set, and the papers published in the five subsequent years (from 1998 to 2002) as the test set for evaluations. Table 3 illustrates the considered journals in this dataset along with the number of extracted papers from each journal and the size of the training-set and the test-set.

Our prepared dataset includes the following information for each paper: Paper identifier (a number between one and 175,432), journal identifier, publication year, and the citation counts from the 0th to 14th year after publication (c_0 to c_{14} citation count values). Actually we set $n = 14$ (according to Table 1, n is the last year after publication in which the citation count is predicted). Table 4 illustrates a small window of the dataset for five sample papers. As it can be seen from the samples, extracting a simple pattern of citations is not trivial, and an intelligent method is necessary for citation pattern prediction.

Table 3
The selected journals, the number of extracted papers, and the size of training and test sets in our experiments. Data are gathered from Web of Science repository.

Journal name	Paper count in the dataset	Training-set size	Test-set size
Nature	72,797	58,068	14,729
Science	52,646	39,616	13,030
NEJM	39,022	31,081	7941
Cell	10,114	8251	1863
PNAS	853	468	385
Total	175,432	137,484	37,948

Table 4
Features (citations history) of five sampled papers.

Paper ID	Journal ID	Publication year	c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}	c_{12}	c_{13}	c_{14}
P1	NATURE	1999	1	19	21	22	19	20	17	11	15	13	12	13	14	9	10
P2	NATURE	1989	0	1	2	0	0	0	0	0	0	0	0	0	0	0	0
P3	NEJM	1980	0	3	2	3	2	1	1	1	0	0	0	0	0	0	0
P4	CELL	1994	2	20	29	25	16	23	14	9	9	7	8	4	5	6	9
P5	SCIENCE	1983	0	5	8	11	16	6	2	6	6	1	4	1	1	0	2

5. Results

In this section, we illustrate the results of our comprehensive experiments, empirical evaluations, and a comparison to the baselines.

5.1. Measurement criteria

We utilized two popular criteria in order to evaluate the proposed method and compare it with the baselines. First, root mean square error (RMSE) and second, the coefficient of determination (R^2). RMSE measures the variation of the predicted values to the actual values and thus, lower values of RMSE are desirable. R^2-score measures the correlation between the actual and the predicted values. The R^2 value is always less than or equal to 1, where $R^2=0$ means no correlation and $R^2=1$ shows a perfect correlation between the predicted and actual values and therefore, higher values of R^2 are desirable. Eqs. (4) and (5) illustrate RMSE and R^2 measurements respectively, in which y is the set of actual values, $\hat{y}$ is the set of the predicted values, and $\bar{y}$ is the average of y_i values.

$$RMSE(y,\hat{y}) = \sqrt{\frac{1}{N_{samples}} \sum_{i=1}^{N_{samples}} (y_i - \hat{y}_i)^2} \tag{4}$$

$$R^2(y,\hat{y}) = 1 - \frac{\sum_{i=1}^{N_{samples}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_{samples}} (y_i - \bar{y})^2} \tag{5}$$

5.2. Evaluation results

We set up different experiments in order to evaluate the proposed method. According to the training-set and the test-set (described in Section 4), we computed the accuracy of the proposed method and the baselines according to two criteria of RMSE and R^2.

As the first motivating experiment, we employed the proposed method and the baselines in order to predict the citation count of 15 randomly selected papers among the 100 highly cited papers of our dataset. Fig. 3 illustrates the actual citation counts (the ground truth) along with the predictions of different methods when $k=5$, i.e., the citation count history up to the fifth year after publication is utilized as the input of the prediction algorithms. As the figures show, there is no simple and trivial pattern of citations in different sampled papers and therefore, predicting the future citations of a paper is not an easy task. Moreover, the proposed method outperforms the baselines (NNCP is usually the most close line to the ground-truth) for most of the sampled papers. After this motivating experiment, comprehensive and quantitative experiments are following.

In the second experiment, we set $n=14$ and $k=5$ (refer to Table 1) which means the citation counts up to the fifth year after publication are used to predict the citation count until the 14th year after publication. In this experiment, the accuracy of the methods are measured in two modes: First, the accuracy of $\hat{C}$ and second, the average accuracy of $\hat{c}_i$ values. In other words, the average accuracy of the yearly predictions is measured ($\hat{c}_i$) along with the aggregated accuracy of the predicted values ($\hat{C}$). Fig. 4 illustrates the evaluation results of this experiment. As the figure shows, the proposed method (named NNCP) results in higher values of R^2 in both yearly and aggregated (total) modes, and this fact is consistent for all of the

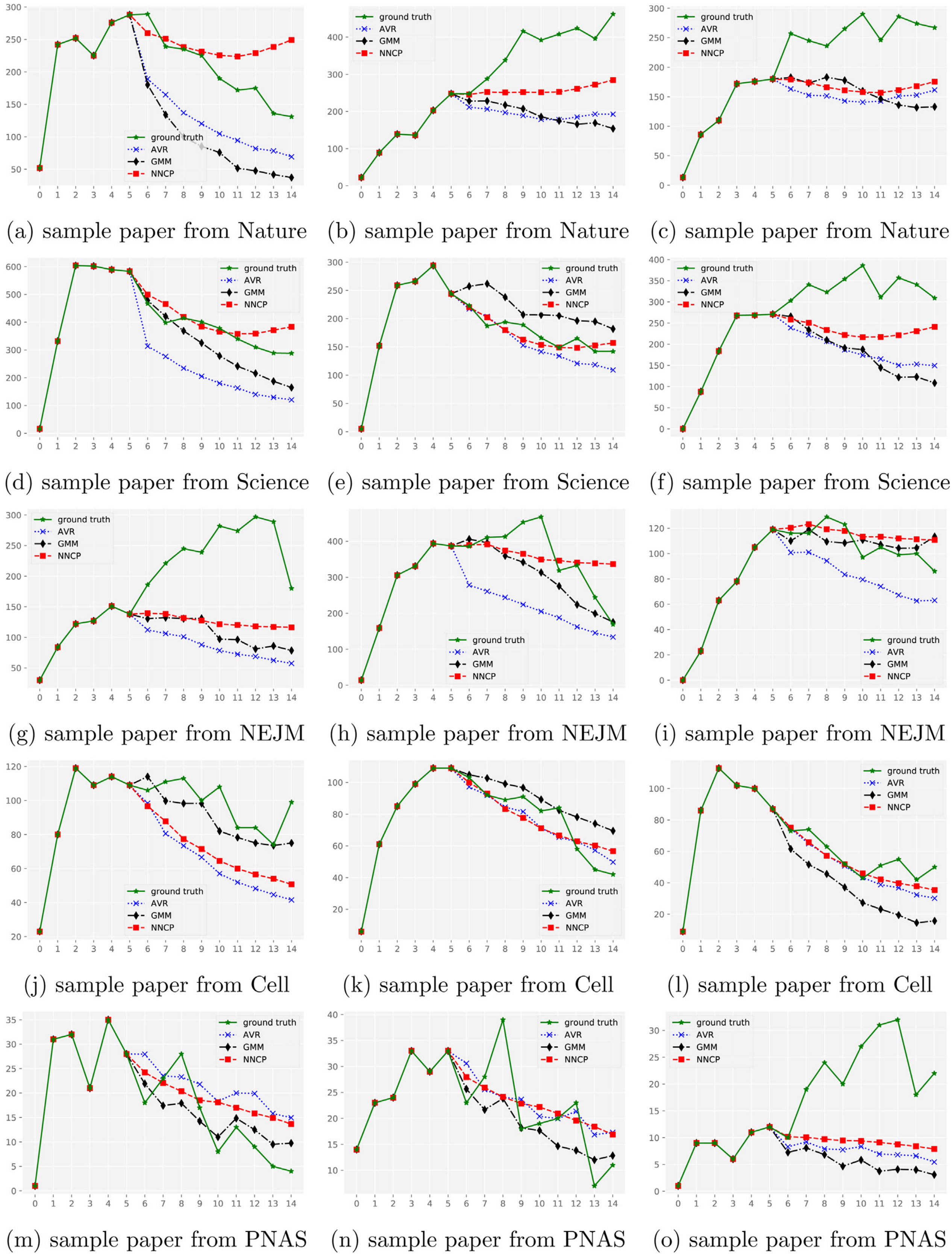
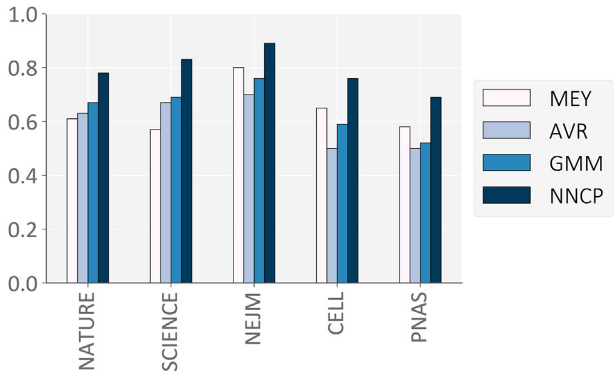
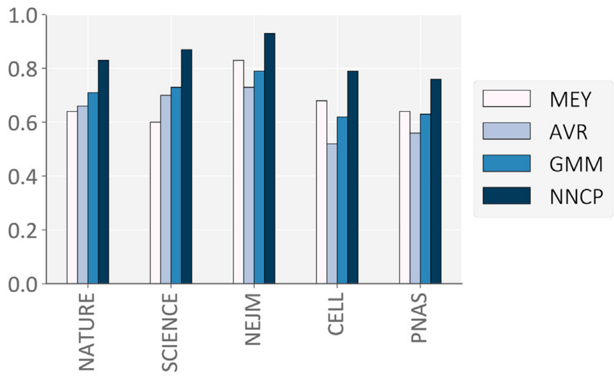


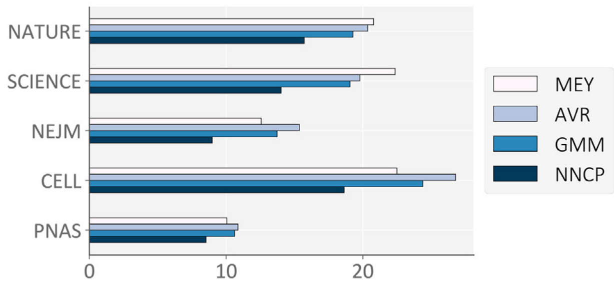
Fig. 3. Comparison of different methods for citation count prediction of 15 randomly sampled papers.



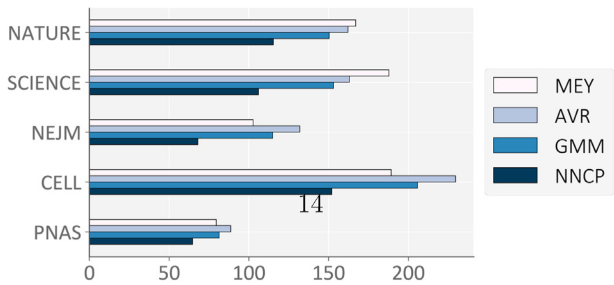
(a) Average yearly R^2 score



(b) Total R^2 score

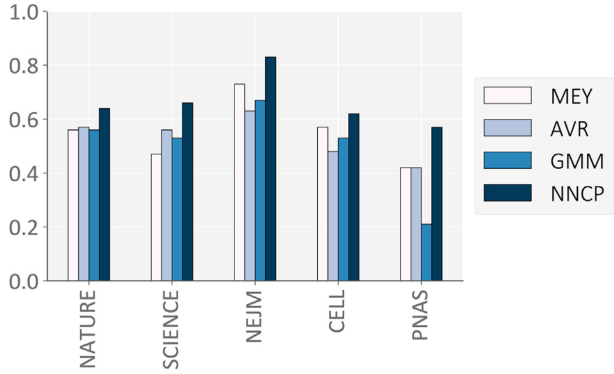


(c) Average yearly RMSE

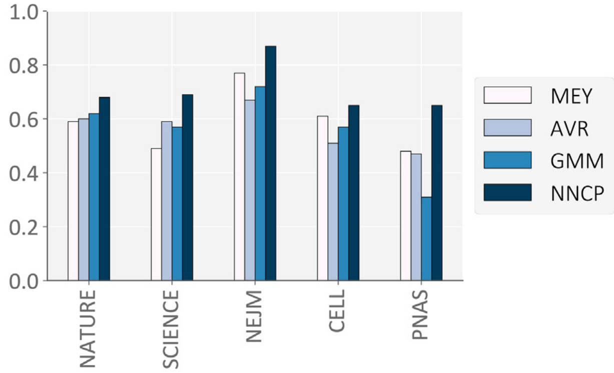


(d) Total RMSE

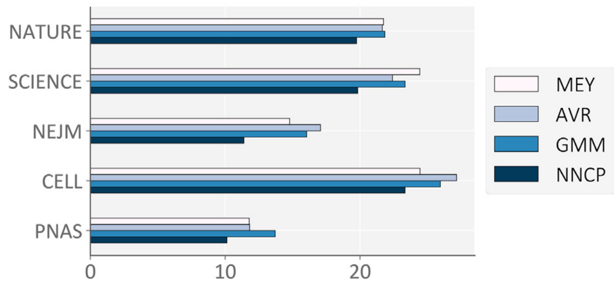
Fig. 4. Evaluation results for proposed method and the baselines in $k = 5$ mode.



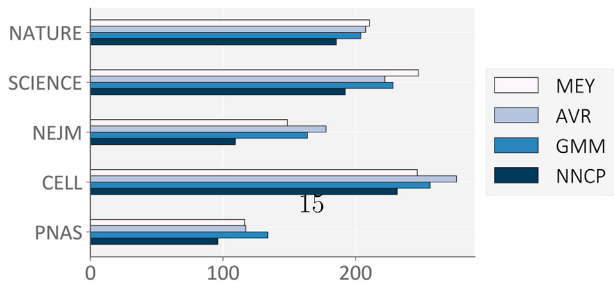
(a) Average yearly R^2 score



(b) Total R^2 score



(c) Average yearly RMSE



(d) Total RMSE

Fig. 5. Evaluation results for proposed method and the baselines in $k = 3$ mode.

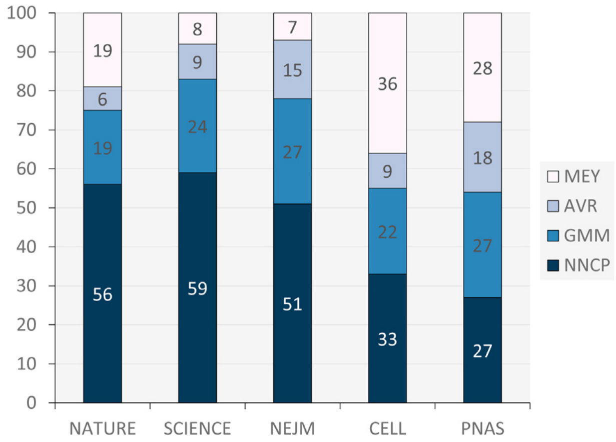


Fig. 6. Frequency of best prediction score for the 100 highly cited papers when $k = 5$.

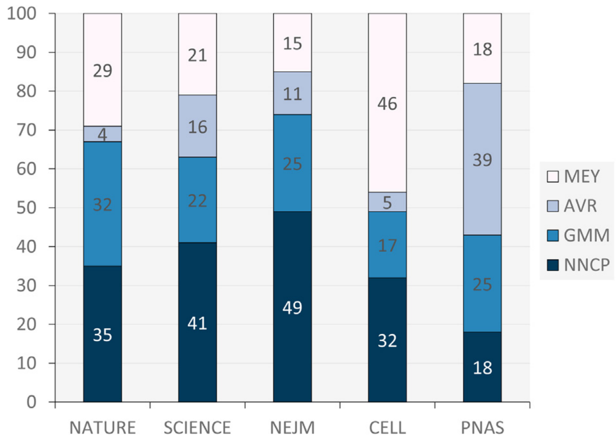
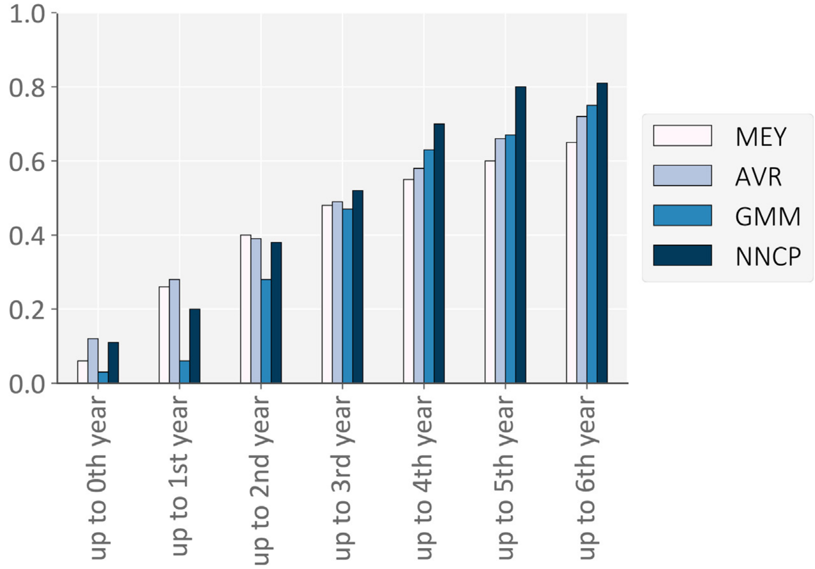


Fig. 7. Frequency of best prediction score for the 100 highly cited papers when $k = 3$.

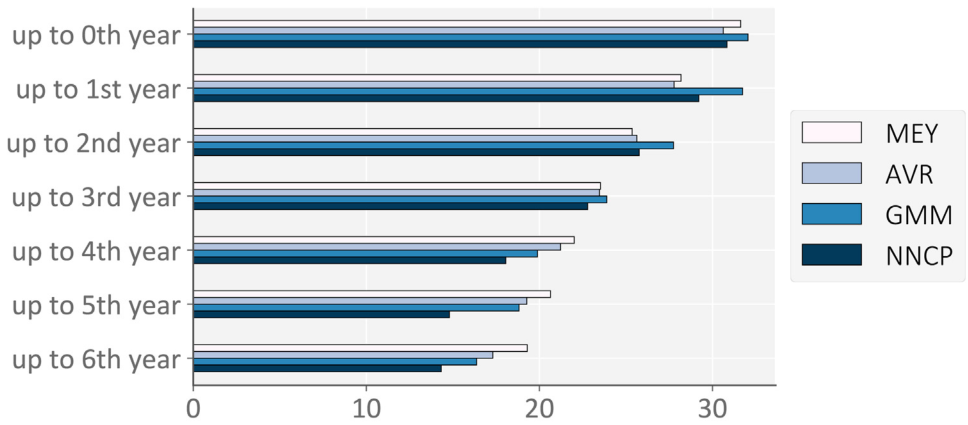
considered journals. Moreover, the proposed method results in less RMSE value for all the journals in both yearly and total modes. Therefore, the proposed method outperforms all the baseline methods in $k = 5$ for all of the considered journals.

In the third experiment, we limit the input information to the citation count up to the third year of publication, and we repeat the second experiment but for $k = 3$. In this experiment, only the citation history up to the third year after publication is used in order to predict the citation counts in the rest of the subsequent years until the 14th year of publication. Fig. 5 shows the evaluation results of this experiment. As the figure shows, the proposed method still outperforms all the baselines in the $k = 3$ mode according to both RMSE and R^2 criteria for all journals of the dataset. It is also worth noting that the prediction results of all the methods are more accurate in the $k = 5$ than those of the $k = 3$ mode according to both RMSE and R^2 scores, because with $k = 5$ more input information is available for the methods (citation counts of 5 years after publication, instead of only three years citation history) in order to investigate the citation counts in the future.

One of the important applications of citation count prediction methods is to detect highly cited papers in advance. Some efforts exist in the literature for early detection of highly cited papers (Newman, 2014; Sarigöl et al., 2014). Therefore, in the next experiment, we evaluate the ability of the proposed method and the baselines regarding their accuracy in predicting citations of highly cited papers. We considered top 100 highly cited papers of each journal in our dataset published between 1998 and 2002 according to their total citations up to 2018. Then, we utilized the proposed method and the baselines in order to predict the citation count of those highly cited papers. We ran the experiment in two phases of $k = 3$ and $k = 5$, and we computed the error of the methods according to the RMSE criterion. Finally, we ranked different methods according to their accuracy of citation prediction. Figs. 6 and 7 show the fraction of samples (highly cited papers) that each method has resulted the best prediction for $k = 5$ and $k = 3$ respectively. For example, the first bar of the Fig. 6 shows that the proposed method (NNCP) is much more accurate than the baselines when they are employed to predict the citation count of the 100 highly cited papers of Nature. In other words, this bar shows that our proposed method results in the best citation count prediction accuracy among the baselines (in 56 out of 100 samples) for the highly cited papers of Nature. Fig. 7 shows the result of a similar experiment but for $k = 3$ (when only the citation count up to the third year after publication is utilized as the input). As Figs. 6 and 7 show, in most cases, the proposed method outperforms the baseline methods in predicting the



(a) R^2 score



(b) RMSE

Fig. 8. Sensitivity analysis of different methods to parameter k. The figure shows the prediction score of different methods for predicting citations of 7th to 14th years of publication, when only the citation count of the first $k+1$ years are available, for $0 \le k \le 7$.

citation count of the highly cited papers, and this fact is valid in both $k=5$ and $k=3$ cases. But in the case of PNAS and Cell journals, our proposed method is not ranked first among the baselines. This is mainly because Cell and PNAS include fewer samples in our dataset (fewer number of papers). In other words, while Figs. 4 and 5 showed that our proposed method outperforms all the baselines when we consider all the papers, Figs. 6 and 7 show that in the case of highly-cited papers, our proposed method is more effective than the baselines when we have enough data in the experiments (e.g., in the case of Nature, Science, and NEJM, but not for Cell and PNAS). This fact is also consistent with the results of the sampled papers illustrated in Fig. 3.

5.3. Sensitivity analysis

As the final experiment, we analyzed the overall accuracy of the considered methods according to different values of k ($0 \le k \le 7$). In this experiment, all the papers of the dataset published in the five journals within 1980–1997 are considered as the training set, and all the papers of the dataset published from 1998 to 2002 are used as the test set. Therefore, this experiment also includes the largest training set and test set utilized in this research. In this evaluation scenario, different methods are utilized in order to predict the citation count of the papers in the test set from the 7th to 14th year after the publication. In other words, the methods are employed to predict c_7, c_8, ..., c_{14} values. Fig. 8 shows the overall accuracy of different methods according to the R^2 score and RMSE with respect to different values of k. When $k=0$ the citation count of the papers only in the publication year is utilized as the input of the methods, and when $k = 6$, citation count history of the publication year until the 6 year after publication is utilized. As the figure shows, by increasing k, more information about the citation history of the papers in the early years of publication is fed to the methods and therefore, the accuracy of different methods improves, as expected. Moreover, the proposed method can outperform baselines when the citation history of the papers is available at least for the first three years (when $k \geq 3$). Therefore, when sufficient information about citation history of the papers is obtainable, the proposed method outperforms the baselines.

5.4. Discussion

We reviewed several experiments which show effectiveness of the proposed method when it compared with the baseline methods. Among the baselines, MEY is a naïve prediction method, but its accuracy is considerable in some limited situations. For example, when we have a small dataset with limited number of sample data and we want to predict the citation count of highly cited papers, MEY method is effective in comparison to other examined methods (see the case of Cell and PNAS journals in Figs. 6 and 7). MEY baseline helps investigate situations in which basic heuristics are effective in citation prediction and moreover, it serves as a simple baseline to evaluate other more complicated methods.

AVR and GMM baselines, follow a local modeling technique called lazy learning in which no model is actually learned from the training data, but the training data are memorized instead. One disadvantage of lazy learning techniques, such as the cases of AVR and GMM, is the need to a large space for storing the entire training dataset. Moreover, noisy training data may decrease the accuracy of such methods. Additionally, AVR and GMM methods actually utilize the citation pattern of only a few (similar) existing papers in order to predict the citation count of the target paper, while NNCP makes use of the whole dataset of papers in order to learn the citation pattern of the scientific papers. It is also worth noting that lazy learning methods are usually slower than their eager learning counterparts. AVR and GMM methods have to compare the input data (the citation pattern of the target paper in its early publication years) with many samples of the dataset in order to find the most similar papers, and this process is repeated for each new query (new target paper). On the other hand, our proposed method once learns a prediction model from the training data and then, applying the model for a new data sample (a new target paper) will be an easy and fast task. In other words, NNCP is considerably faster than AVR and GMM methods in predicting the citation counts, because NNCP has passed a training phase and it has learned a prediction model which is needless to re-examine a large set of training data. Therefore NNCP outperforms AVR and GMM with respect to both efficiency and accuracy.

6. Conclusion

In this paper, we proposed a novel method for citation count prediction, which is based on artificial neural networks. We employed modern deep learning techniques (such as RNNs and sequence-to-sequence model) in order to learn a prediction method based on the sequence pattern of the citations from early years of publication of a paper. The comprehensive evaluations show that the proposed method outperforms state-of-the-art methods of citation count prediction with respect to the accuracy of the prediction and the ability to predict the citations of highly-cited papers. Many challenges arise when applying deep learning to the problem of citation prediction. Those challenges include: Mapping the citation problem into a deep learning problem, designing the learning and experimentation scenarios for this specific problem, choosing the evaluation criteria, selecting the best fitting deep learning architecture for this specific application, and tuning the hyperparameters of the selected neural network architecture.

As the future works, we will include more features as the inputs of the prediction algorithm and deeper layers as the hidden layers of the proposed neural network. We will also investigate applying the proposed method for higher-level tasks, such as predicting highly cited papers and authors, and estimating the h-index of the authors in the future.

Author contributions

Ali Abrishami: Conceived and designed the analysis, Collected the data, Contributed data or analysis tools, Performed the analysis.

Sadegh Aliakbary: Conceived and designed the analysis, Performed the analysis, Wrote the paper.

References

Acuna, D. E., Allesina, S., & Kording, K. P. (2012). Future impact: Predicting scientific success. *Nature*, *489*(7415), 201–202. ISSN 0028-0836

Amjad, T., Ding, Y., Xu, J., Zhang, C., Daud, A., Tang, J., et al. (2017). Standing on the shoulders of giants. *Journal of Informetrics*, *11*(1), 307–323. ISSN 1751-1577.

Ayaz, S., Masood, N., & Islam, A. M. (2018 Mar). Predicting scientific impact based on h-index. *Scientometrics*, *114*(3), 993–1010. ISSN 1588-2861.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). *Neural machine translation by jointly learning to align and translate*. CoRR, abs/1409.0473.

Bornmann, L., Leydesdorff, L., & Wang, J. (2014). How to improve the prediction based on citation impact percentiles for years shortly after the publication date? *Journal of Informetrics*, *8*(1), 175–180. ISSN 1751-1577.

Bütün, E., Kaya, M., & Alhajj, R. (2017). A supervised learning method for prediction citation count of scientists in citation networks. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ASONAM '17* (pp. 952–958). ISBN 978-1-4503-4993-2.

Cao, X., Chen, Y., & Ray Liu, K. J. (2016). A data analytic approach to quantifying scientific impact. *Journal of Informetrics*, *10*(2), 471–484. ISSN 1751-1577.

Castillo, C., Donato, D., & Gionis, A. (2007). Estimating number of citations using author reputation. In *String Processing and Information Retrieval*. pp. 107–117. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-540-75530-2.
Chan, H. F., Mixon, F. G., & Torgler, B. (2018 Mar). Relation of early career performance and recognition to the probability of winning the Nobel prize in economics. *Scientometrics*, *114*(3), 1069–1086. ISSN 1588-2861.
Cho, K., van Merrienboer, B., G“ulçehre, Ç., Bougares, F., Schwenk, H., & Bengio, Y. (2014). *Learning phrase representations using RNN encoder-decoder for statistical machine translation*. CoRR, abs/1406.1078.
Chollet, F. (2017). *Deep Learning with Python* (1st edition). Greenwich, CT, USA: Manning Publications Co. ISBN 1617294438, 9781617294433.
Dahl, G. E., Yu, D., Deng, L., & Acero, A. (2012 Jan). Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, *20*(1), 30–42. ISSN 1558-7916.
Daud, A., Ahmed, W., Amjad, T., Nasir, J. A., Aljohani, N. R., Abbasi, R. A., et al. (2017). Who will cite you back? Reciprocal link prediction in citation networks. *Library Hi Tech*, *35*(4), 509–520.
Dong, Y., Johnson, R. A., & Chawla, N. V. (2016 March). Can scientific impact be predicted? *IEEE Transactions on Big Data*, *2*(1), 18–30. http://dx.doi.org/10.1109/TBDATA.2016.2521657. ISSN 2332-7790
Dong, Yuxiao, Johnson, R. A., & Chawla, N. V. (2015). Will this paper increase your h-index?: Scientific impact prediction. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining, WSDM '15*. pp. 149–158. New York, NY, USA: ACM. ISBN 978-1-4503-3317-7.
Fiala, D., & Tutoky, G. (2017). Pagerank-based prediction of award-winning researchers and the impact of citations. *Journal of Informetrics*, *11*(4), 1044–1068. ISSN 1751-1577.
Eugene, G. (1998). The use of journal impact factors and citation analysis for evaluation of science. In *41st Annual Meeting of the Council of Biology Editors*.
Eugene, G. (2006). The history and meaning of the journal impact factor. *JAMA*, *295*(1), 90–93.
Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*. pp. 315–323. Fort Lauderdale, FL, USA: PMLR.
Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
Graves, A., Mohamed, A., & Hinton, G. (2013 May). Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. pp. 6645–6649.
Havemann, F., & Larsen, B. (2015). Bibliometric indicators of young authors in astrophysics: Can later stars be predicted? *Scientometrics*, *102*(2), 1413–1434. ISSN 1588-2861.
He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A., Jaitly, N., et al. (2012). Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, *29*(6), 82–97. ISSN 1053-5888.
Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, *102*(46), 16569–16572. ISSN 0027-8424.
Karpathy, A., & Fei-Fei, L. (2017). Deep visual-semantic alignments for generating image descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *39*(4), 664–676. ISSN 0162-8828.
Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying sleeping beauties in science. *Proceedings of the National Academy of Sciences*, *112*(24), 7426–7431. ISSN 0027-8424.
Ketcham, Catherine M. (2007). Predicting impact factor one year in advance. *Laboratory Investigation*, *87*(6), 520.
Klimek, P., Jovanovic, A. S., Egloff, R., & Schneider, R. (2016). Successful fish go with the flow: Citation impact prediction based on centrality measures for term-document networks. *Scientometrics*, *107*(3), 1265–1282. ISSN 1588-2861.
Kosteas, Vasilios D. (2018). Predicting long-run citation counts for articles in top economics journals. *Scientometrics*, *115*(3), 1395–1412. ISSN 1588-2861.
Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems 25*. pp. 1097–1105. Curran Associates, Inc.
Lai, S., Xu, L., Liu, K., & Zhao, J. (2015). Recurrent convolutional neural networks for text classification. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, AAAI'15*. pp. 2267–2273. AAAI Press. ISBN 0-262-51129-0.
Lamb, C. T., Gilbert, S. L., & Ford, A. T. (2018). Tweet success? Scientific communication correlates with increased citations in ecology and conservation. *PeerJ*, *6*, e4564. ISSN 2167-8359.
Li, J., & Ye, F. Y. (2016). Distinguishing sleeping beauties in science. *Scientometrics*, *108*(2), 821–828. ISSN 1588-2861.
Mazloumian, A. (2012 11). Predicting scholars' scientific impact. *PLOS ONE*, *7*(11), 1–5.
McNamara, D., Wong, P., Christen, P., & Ng, K. S. (2013). Predicting high impact academic papers using citation network features. In *Trends and Applications in Knowledge Discovery and Data Mining*. pp. 14–25. Berlin, Heidelberg: Springer Berlin Heidelberg. ISBN 978-3-642-40319-4.
Mikolov, T., Karafiát, M., Burget, L., Černocký, J., & Khudanpur, S. (2010). Recurrent neural network based language model. *Eleventh Annual Conference of the International Speech Communication Association*.
Moed, H. F. (2005). *Citation analysis in research evaluation (information science and knowledge management)*. Berlin, Heidelberg: Springer-Verlag. ISBN 1402037139.
Moed, H. F., Colledge, L., Reedijk, J., Moya-Anegon, F., Guerrero-Bote, V., Plume, A., et al. (2012). Citation-based metrics are appropriate tools in journal assessment provided that they are accurate and used in an informed way. *Scientometrics*, *92*(2), 367–376. ISSN 1588-2861.
Newman, M. E. J. (2014). Prediction of highly cited papers. *Europhysics Letters*, *105*(2), 28002.
Oppenheim, C. (1995). The correlation between citation counts and the 1992 research assessment exercise ratings for British library and information science university departments. *Journal of Documentation*, *51*(1), 18–27.
Pobiedina, N., & Ichise, R. (2016). Citation count prediction as a link prediction problem. *Applied Intelligence*, *44*(2), 252–268. ISSN 1573-7497.
Revesz, P. Z. (2014). Data mining citation databases: A new index measure that predicts Nobel prizewinners. In *Proceedings of the 19th International Database Engineering & Applications Symposium, IDEAS '15*. pp. 1–9. New York, NY, USA: ACM. ISBN 978-1-4503-3414-3.
Silva, M. R. (2016). Journal impact factors for the year-after the next can be objectively predicted. *Medical Express*, *3*(5).
Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533.
Sarigöl, E., Pfitzner, R., Scholtes, I., Garas, A., & Schweitzer, F. (2014). Predicting scientific success based on coauthorship networks. *EPJ Data Science*, *3*(1), 9. ISSN 2193-1127.
Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), CVPR '12*. pp. 3642–3649. Washington, DC, USA: IEEE Computer Society. ISBN 978-1-4673-1226-4.
Severyn, A., & Moschitti, A. (2015a]). Learning to rank short text pairs with convolutional deep neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '15*. pp. 373–382. New York, NY, USA: ACM. ISBN 978-1-4503-3621-5.
Severyn, A., & Moschitti, A. (2015b]). Twitter sentiment analysis with deep convolutional neural networks. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '15*. pp. 959–962. New York, NY, USA: ACM. ISBN 978-1-4503-3621-5.
Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition*. CoRR, abs/1409.1556.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958.
Sutskever, I., Martens, J., & Hinton, G. (2011). Generating text with recurrent neural networks. In *Proceedings of the 28th International Conference on International Conference on Machine Learning, ICML'11*. pp. 1017–1024. USA: Omnipress. ISBN 978-1-4503-0619-5.
Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems 27*. pp. 3104–3112. Curran Associates, Inc.

van Raan, A. F. J. (2004). Sleeping beauties in science. *Scientometrics*, *59*(3), 467–472. ISSN 1588-2861.
Wan, J., Wang, D., Hoi, S. C. H., Wu, P., Zhu, J., Zhang, Y., et al. (2014). Deep learning for content-based image retrieval: A comprehensive study. In *Proceedings of the 22Nd ACM International Conference on Multimedia, MM '14*. pp. 157–166. New York, NY, USA: ACM. ISBN 978-1-4503-3063-3.
Wang, D., Song, C., & Barabási, A. (2013). Quantifying long-term scientific impact. *Science*, *342*(6154), 127–132. ISSN 0036-8075.
Werbos, P. J. (1990). Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE*, *78*(10), 1550–1560. ISSN 0018-9219.
Wildgaard, L., Schneider, J. W., & Larsen, B. (2014). A review of the characteristics of 108 author-level bibliometric indicators. *Scientometrics*, *101*(1), 125–158. ISSN 1588-2861.
Wu, X.-f., Fu, Q., & Rousseau, R. (2008). On indexing in the web of science and predicting journal impact factor. *Journal of Zhejiang University Science B*, *9*(7), 582–590. ISSN 1862-1783.
Yan, R., Tang, J., Liu, X., Shan, D., & Li, X. (2011). Citation count prediction: Learning to estimate future citations for literature. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management, CIKM '11*. pp. 1247–1252. New York, NY, USA: ACM. ISBN 978-1-4503-0717-8.
Yan, R., Huang, C., Tang, J., Zhang, Y., & Li, X. (2012). To better stand on the shoulder of giants. In *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '12*. pp. 51–60. New York, NY, USA: ACM. ISBN 978-1-4503-1154-0.
Yu, T., Yu, G., Li, P.-Y., & Wang, L. (2014). Citation impact prediction for scientific papers using stepwise regression analysis. *Scientometrics*, *101*(2), 1233–1252. ISSN 1588-2861.